iSCSI technology:

A convergence of networking and storage

technology brief

Abstract

This technology brief introduces readers to the Internet SCSI (iSCSI) protocol; places iSCSI in the context of existing I/O fabrics such as Fibre Channel and IP; discusses the initial applications of IP storage technologies; and addresses the future impact of Remote Direct Memory Access technology.

Introduction

As businesses are flooded by an endless stream of data, they can choose from three major storage approaches: direct-attached storage (DAS), network-attached storage (NAS), and storage area networks (SANs). DAS describes all internal and external storage devices that are directly connected to a host computer, typically using parallel Small Computer Systems Interface (SCSI). DAS provides a very attractive price/performance ratio and accounts for roughly 80 percent of storage sold today. The NAS model consists of devices that are connected directly to a local area network (LAN). NAS systems allow multiple clients running different operating systems to access data using network file protocols. Currently, Gigabit Ethernet NAS infrastructures are being deployed in Information Technology (IT) environments. A SAN is a dedicated network of storage systems that provide storage capabilities to one or more servers over a high-speed interconnect, usually Fibre Channel (FC). FC SANs create flexible, consolidated, and manageable storage environments.

NAS systems differ from DAS and SANs in how they transport data. NAS typically moves data in the form of files over Internet Protocol (IP)-based networks. NAS uses host-friendly file semantics[1], such as Network File System (NFS) and Common Internet File System (CIFS), to store and retrieve files in their native format. NAS-based storage solutions leverage LANs, allowing IT managers to use their existing Ethernet/IP knowledge base and network management tools. DAS and SANs transport data using SCSI protocols (block-level I/O), which is more efficient than file-level I/O for moving some kinds of data between hosts and storage.

New transport protocols are available for moving block-level data over IP networks. With the advent of block data transport over IP, all storage solutions will some day be able to leverage Ethernet infrastructures. HP is developing IP-based storage solutions that will leverage customers' existing Ethernet knowledge and infrastructure. HP's IP-based storage solutions will incorporate iSCSI (SCSI protocol encapsulated in IP), FC over IP (FCIP), and Remote Direct Memory Access (RDMA) technologies that integrate seamlessly and complementarily with existing DAS, NAS, and SANs. This technology brief describes various IP storage technologies but focuses on iSCSI technology and its role in the HP Adaptive Infrastructure.

The changing storage environment

Businesses have had to manage increasing numbers of individual DAS devices by protecting data with backup, replication, disaster recovery, and other techniques. The resulting management complexities have driven organizations to deploy NAS systems and SANs. These networked storage solutions reduce the overall expense associated with data management and, therefore, are expected to become the dominant paradigm in enterprises.

NAS systems have become an appealing storage solution due to the pervasiveness and knowledge of IP and file-based management. They also offer a relatively low cost for life cycle data management when value-added features like "snap shot" and cloning are added to the solution.

[1] NAS uses file transfer protocols such as Network File System (NFS) for UNIX®/Linux and Common Internet File System (CIFS) for Windows® to translate between an application's file I/O request and the native file system on the NAS server.

Fibre Channel-based SANs have become the dominant paradigm for large data stores and clustered server deployments because they offer large-scale consolidation, performance, and manageability benefits. Additional drivers of SAN adoption are the continued decrease in Fibre Channel infrastructures costs and the increase in interoperability and manageability. Still, there are additional consolidation and management opportunities that will allow Ethernet-based SANs to complement FC SANs. An example of this is server blade deployments where Ethernet is integrated into the infrastructure and where applications do not require performance or availability beyond what an industry-standard network interface controller (NIC) and iSCSI software driver can provide.

IT managers demand increasing storage management ease and flexibility, more reliable remote data access, and remote management of their IT environment from a single management interface (portal). In other words, customers are seeking adaptive storage infrastructures that are accessible on demand.

The HP Adaptive Infrastructure will leverage industry-standard IP-based storage technologies such as iSCSI and FCIP. IP-based storage technologies will enable IT managers to create and manage heterogeneous environments where NAS and FC SANs can be integrated over a common network.

iSCSI technology—the focus of this paper— enables dynamic resource optimization to be extended beyond today's Fibre Channel environment because it leverages a pervasive IP fabric infrastructure to address the block data requirements of today’s enterprise storage applications.

iSCSI technology

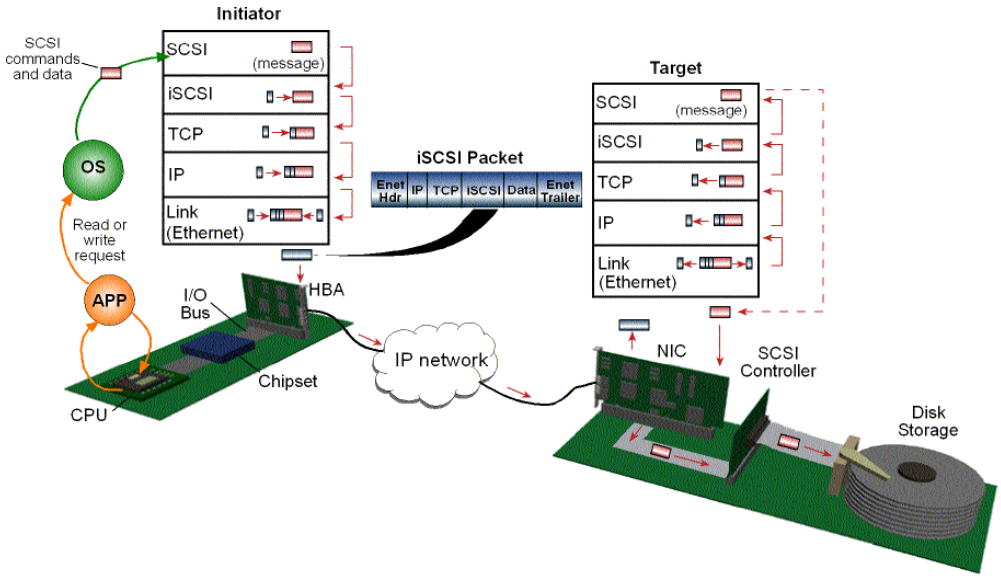
The iSCSI protocol defines the rules and processes for transporting SCSI (block-level) data over a TCP/IP network. The iSCSI standard was developed by the Internet Engineering Task Force (IETF). HP is working with the IETF and other standards organizations to ensure that IP-based storage standards are optimized for future needs of customers.

How iSCSI works

iSCSI follows the SCSI architectural model, which is based on message exchange between an initiator and a target. In the iSCSI model, initiators and targets are identified by a unique SCSI device name. Because iSCSI transport occurs over a network fabric instead of a direct cable connection, the initiator and target have multiple IP addresses associated with their iSCSI names.

Figure 1 illustrates a message exchange between an initiator and a target. The process begins when an application sends a request to the operating system (OS) to read or write data. The OS generates the appropriate SCSI commands and data request in the form of a message. Before the message can be sent over an IP network, it is processed through iSCSI to encapsulate the request into the TCP/IP protocol stack (attaching routing, error checking, and control information) for transmission over the network. This can be accomplished using driver- or OS-level software, or it can be offloaded to the Host Bus Adapter (HBA) (see "iSCSI technical challenges"). The HBA transmits the packets over the IP network. When the packets reach the target device, they go through a reverse process to reassemble (sequence) the data, which is then moved to the SCSI controller. The SCSI controller fulfills the request by writing data to or reading data from the target device. If it is a read transaction, the target returns data to the initiator using the iSCSI protocol.

Figure 1. Message exchange between an initiator and target using the iSCSI protocol mode



iSCSI technical challenges

One major challenge concerning iSCSI technology is its performance. iSCSI processing functionality can run in host driver software and be sent over a standard Ethernet NIC, or it can be optimized in hardware for better performance on an iSCSI HBA. If TCP/IP and iSCSI protocol processing are performed in software, the amount of CPU utilization devoted to non-application tasks increases significantly. With 1-Gb/s Ethernet, for example, protocol processing can increase the utilization of a 1-GHz CPU to 100 percent. Consequently, it is desirable to offload protocol stack processing to a dedicated hardware component— the adapter card (HBA or NIC)—to reduce CPU utilization to acceptable levels. Also, this sort of offloading can greatly improve the speed with which requests are processed and sent to the network.

TCP/IP and iSCSI protocol processing can be accomplished on a network adapter using offload engines. A TCP/IP offload engine (TOE) is software and/or hardware embedded in a chip or on a network adapter that performs TCP/IP stack processing. Similarly, an iSCSI offload engine is software and/or hardware embedded in a network adapter to perform iSCSI protocol processing. Dual-purpose (networking and storage) iSCSI adapters are available to perform both TOE and iSCSI offload functions. iSCSI adapters (HBAs and NICs) are currently at 1 Gb/s; 10-Gb/s adapters are expected to start arriving in 2003.

HP expects iSCSI performance to improve with the use of TCP/IP offload engines, more efficient IP stacks, and faster Ethernet fabrics. iSCSI will co-exist with Fibre Channel, giving each enterprise a broad choice of networked storage solutions for many applications.

iSCSI benefits

iSCSI enables improvements in the economics, operating distance, and manageability of storage networks. iSCSI also leverages the security capabilities of IP networks. These benefits are described below.

Cost

iSCSI can potentially achieve a lower total cost of ownership (TCO) than Fibre Channel. Depending on application demands, Ethernet SANs can leverage existing network infrastructures. Although the

initial cost of iSCSI adapters may be comparable to that of FC host bus adapters, eventual industry-wide acceptance and volume production are expected to result in a lower price for iSCSI adapters.

Operating distance

It is possible for an iSCSI-based network to economically span great distances using commonly available WANs. Longer operating distances will allow customers to mirror and archive data at remote sites for disaster protection.

Manageability

Using bridging products, iSCSI allows customers to present SAN capacity over an IP network. iSCSI can use some existing IP-based network management software. Note that iSCSI devices will require storage management tools.

Security

IP networks have a well-defined security infrastructure (encryption and authentication) that makes iSCSI viable for remote back up and disaster recovery applications. Fibre Channel networks are primarily protected with physical security.

Encryption

An iSCSI transfer can optionally encrypt each packet, ensuring security until the packet is decrypted by the receiver. A set of protocols called IPSEC (developed by the IETF) describes two encryption modes: transport and tunnel. Transport mode encrypts only the data portion (payload) of each packet, but leaves the header untouched. Tunnel mode encrypts both the header and the payload for increased security.

Authentication

iSCSI has provisions to mutually authenticate servers with storage at login. iSCSI uses the Challenge Handshake Authentication Protocol with Diffie-Hellman key protocol (DH CHAP). IPSEC can also provide further protection with per packet authentication.

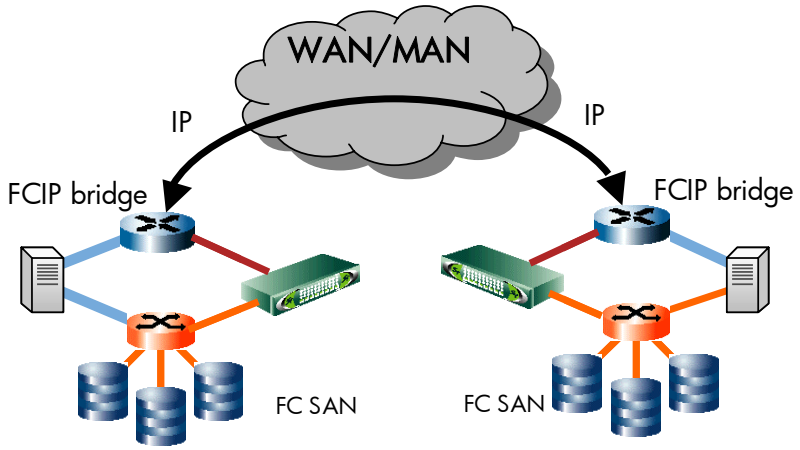
Initial applications for IP storage

IP-based storage solutions will extend access to legacy and future storage resources across local, metropolitan, and wide area networks. HP's IP storage solutions incorporate iSCSI and FCIP, both of which were developed by the IETF IP Storage Working Group. The functions of iSCSI and FCIP differ, but they commonly provide block-level access to storage resources over an IP network. The iSCSI protocol is intended to consolidate DAS, NAS, and FC SAN resources into a single storage network. If a customer already has FC SANs in place, FCIP can be used to connect these SANs across geographically dispersed enterprises. The first opportunities for customers to take advantage of iSCSI and FCIP solutions are described in more detail below.

SAN-to-SAN connection

Some enterprises use multiple FC SANs at different sites that are not part of a single storage network. Linking these geographically separated FC SAN islands is becoming more critical for data protection and accessibility across the enterprise. FC SANs can presently be interconnected up to 10 km (6 miles) using Fibre Channel; however, FCIP can be used to interconnect SANs over much longer distances (Figure 2). FCIP encapsulates FC protocol data into IP packets and tunnels within TCP/IP. This capability allows native FC SANs to communicate with each other across IP networks (Ethernet, SONET, and ATM) with all FC services remaining intact. FCIP requires gateway devices (bridges) that translate between FC and FCIP protocols.

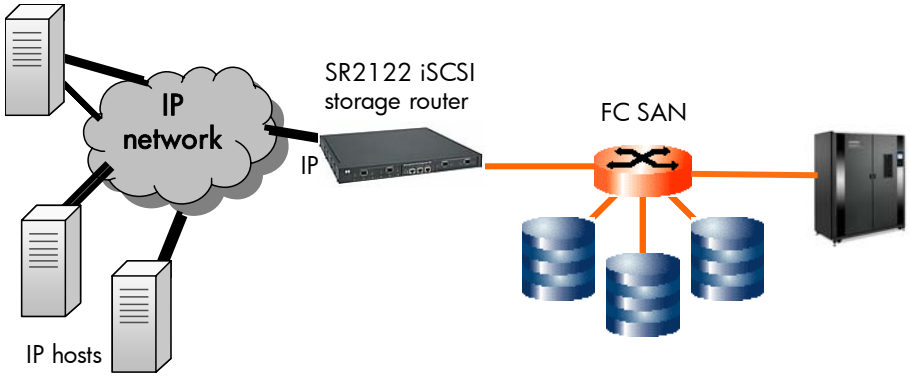
Figure 2. FCIP bridges tunnel through TCP/IP so that FC SANs can communicate.



SAN extension to IP host servers

As shown in Figure 3, iSCSI will also allow FC SAN architectures to be accessed from across IP networks using an iSCSI-to-FC router. For example, the HP StorageWorks SR2122 iSCSI Storage Router provides an excellent option for allowing less critical servers to access existing SAN storage resources with a minimum incremental investment. The SR2122 also leverages existing IT staff expertise for increased productivity. Linking servers with the FC SAN without FC HBAs will become more common as iSCSI-to-FC routers offer reliable and cost-effective connectivity.

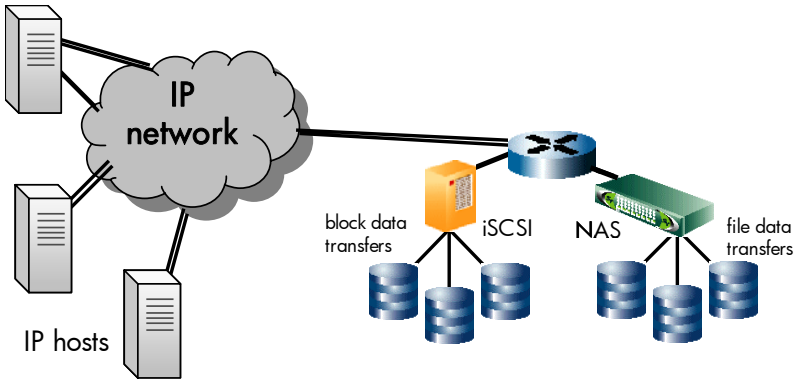
Figure 3. Customers can connect IP host servers to existing FC SANs using an iSCSI to FC router.



Consolidated IP storage

DAS and NAS will continue to satisfy many business needs; however, there is a growing need to consolidate data repositories to simplify data management, backup, and disaster recovery. Like other SAN deployments, native iSCSI storage devices on an IP network will begin to offer the benefits of network storage consolidation with basic functionality. Native iSCSI storage devices will allow data to be stored and accessed anywhere on the network, and storage traffic can be managed through the existing enterprise management tools (Figure 4).

Figure 4. Single IP storage system for converged file and block storage.



Future opportunities for IP storage

Future opportunities for IP storage solutions will be enhanced by the deployment of RDMA and ultimately by the convergence of block and file data management in enterprise environments. RDMA technology was developed to move data from the memory of one computer directly into the memory of another computer with minimal involvement from their CPUs. Additional information included in the RDMA protocol allows a system to place the data directly into its final memory destination without any additional or interim data copies. This "zero copy" or "direct data placement" (DDP) capability provides the most efficient network communication possible between systems.

In the future, it is expected that iSCSI will utilize RDMA hardware to allow zero-copy data placement while dramatically reducing the CPU utilization of iSCSI. Thus, a single RDMA-enabled network adapter can support both low-latency, low CPU-intensive network connectivity and high-performance iSCSI traffic.

Conclusion

With regard to storage networks, iSCSI enables block level data transfer over IP networks to complement existing IP-based file level protocols such as NFS and CIFS. HP will continue to work with the IETF and other standards groups to ensure IP-based storage standards are optimized for the future needs of customers. Finally, HP will continue to invest in technology development and solution delivery that maintains industry leadership and provides effective, robust, and cost effective solutions for customers.

Call to action

To help us better understand and meet your needs for ISS technology information, please send comments about this paper to: TechCom@HP.com.

TC030402TB, 04/2003

Printed in the US